

АЛГЕБРА УЗГОДЖЕННЯ ОНТОЛОГІЧНИХ КОНТЕКСТІВ В УМОВАХ НЕПОВНОЇ ІНФОРМАЦІЇ

Запропоновано алгебру маніпулювання онтологіями, необхідну для формалізації операцій узгодження та пошуку елементів в ієрархіях їх таксономій, представлених різними джерелами даних. Необхідність розроблення такої алгебри мотивована неможливістю створення глобальної узгодженої онтології, тому пропонується проводити узгодження на локальному рівні для тематично пов'язаних інформаційних джерел. Можливими областями застосування розробленої алгебри є системи оцінки інформаційних статей та оптимізації досяжності їх в гіпертекстових мережах.

Досягнення оптимальної структури гіпертекстових посилань певного сегменту мережі Інтернет є однією із задач пошукової оптимізації. Одним із важливих факторів в даній оптимізації є не лише максимізація кількості вхідних та вихідних посилань, а й збереження їх семантичної коректності. Результатом такої оптимізації є підвищення рейтингу цитування Інтернет-ресурсу, його позицій в пошукових системах.

Забезпечення семантичної коректності в мережі WWW є складним процесом, що характеризується великою вартістю, оскільки виконується в ручному режимі спеціалістами з пошукової оптимізації. Необхідно зазначити, що цей процес не стандартизований і виконується, спираючись на досвід спеціалістів, а отже у разі оптимізації певного сегменту мережі різними спеціалістами результат загальної оптимізації є невизначеним. Однак зі стрімким розвитком концепції Semantic Web, що пропонує як стандарти, так і реалізації численних засобів по інтелектуалізації середовища WWW, цей процес можна як стандартизувати, так і автоматизувати.

Основною особливістю Semantic Web є поєднання інформації про предметну область — об'єкти, поняття, відношення та їх відображення на гіпертекстову структуру ресурсів. Це дає можливість перейти від аналізу залежностей гіпертекстових посилань до аналізу предметної області, що описана онтологією, яка лежить в основі Semantic Web.

Існуючі засоби математичної формалізації [1, 3] добре виконують свою функцію у разі здійснення операцій на одиничній онтології, але у випадку декількох різнорідних онтологій немає доступного, цілісного математичного апарату, який би включав як операції по аналізу структури, так і операції з їх узгодження. Отже, у дослідженні та оптимізації контекстів декількох Інтернет-ресурсів необхідним стає узгодження їх онтологій, що фактично означає їх злиття і робить можливим проведення аналізу взаємозв'язків між ними.

Хоча узгодження онтологічних контекстів інформаційних статей, навіть за умови їх належності до схожих предметних областей, є проблематичним, оскільки має місце невизначеність понять [1], пропонується застосувати низку засобів для їх попередньої інтеграції. Оскільки задача узгодження нетривіальна і в основному полуавтоматична, тому важливим є чітка математична формалізація операцій, які необхідно виконати для отримання результату і покращення можливостей до її автоматизації.

В зв'язку із зазначеними складностями, в роботі розв'язується задача математичної формалізації операцій для узгодження онтологій, разом з операцією ідентифікації ієрархій понять з різних джерел для подальшого аналізу і оцінювання гіпертекстової структури на основі семантичної інформації.

Нехай у підмережі ресурсів, що розглядається, маємо множину ресурсів $S = (s_1, s_2, \dots, s_N)$, множину сторінок $P = (p_1, p_2, \dots, p_M)$, множину зв'язків між сторінками $L = (l_1, l_2, \dots, l_K)$, де $l_n = (p_i, p_j)$ та множину онтологій $O = (o_1, o_2, \dots, o_{|P|})$.

Визначимо відношення на них. Відношення ідентичності $(s, p, o) \in R_i$ однозначно ідентифікує інформаційну статтю в мережі ресурсів. Відношення належності сторінок до ресурсів $sR_r p$, фактично, є відношенням еквівалентності на P і задовольняє вимогам рефлексивності, симетричності та транзитивності; $pR_c l$ — відношення, що характеризує зв'язність елементів у підмережі ресурсів; lR_s — визначає множину посилань окремо взятого ресурсу. Добуток відношень $R_r \cdot R_c$ в цьому випадку визначає множину зв'язків сайтів підмережі, що розглядається, в загальному вигляді

$$R_r \cdot R_c = \{(s, l) | s \in S, l \in L \ \& \ (\exists p \in P)(sR_r p \ \& \ pR_c l)\}, \quad (1)$$

тоді

$$R_{ps} = (R_r \cdot R_c | s = S_i, l \in L) \quad (2)$$

визначає множину зв'язків певного Інтернет-ресурсу заданого обмеженням $s = S_i$.

Відношення $oR_s p$ визначає контекст інформаційної статті, а $R_s \cdot R_r$ — онтологічний контекст підмережі Інтернет-ресурсів, тобто

$$U(O) = R_s \cdot R_r = \{(o, s) | o \in O, s \in S \ \& \ (\exists p \in P)(oR_s p \ \& \ pR_r s)\}, \quad (3)$$

тоді

$$R_{osc} = \{R_s \cdot R_r | o \in O, s \in S_i\} \quad (4)$$

визначає онтологічний контекст певного ресурсу.

Формально, онтологія визначена як четвірка $\langle C, R, I, A \rangle$, де C — поняття; R — відношення; I — об'єкти; A — аксіоми. Оскільки онтології реалізуються мовами їх опису, а роботі для цього було обрано RDF, то можемо дещо спростити цей запис. Для представлення онтології будемо використовувати орієнтований маркований граф $O = (V, R)$, де V — вершини графа, що є об'єднанням $V = C \cup I$ та R — відношення між елементами онтології. Фактично, R одночасно виконує дві функції — вказує на зв'язки між поняттями та визначає їх характер, наприклад *rdfs:instanceOf* чи *rdfs:subclassOf*. В такому випадку може бути визначена множина зв'язків

$$E = \left\{ e \mid (n_i, n_j) \in V \ \& \ (n_i R e \ \& \ n_j R e) \right\}, \quad (5)$$

де n_i, n_j — вершини; e — ребро між n_i та n_j .

Іррефлексивне антисиметричне відношення $n_i R_d n_j$ визначає напрямок зв'язку між вершинами, вказуючи на наявність дуги між n_i до n_j .

Визначимо нотації на онтології O . Ребро $e \in E$ має запис (n_i, δ, n_j) , де n_i та n_j — поняття (вершини графа) з V . Функція $\delta = \gamma(e)$ визначає відношення асоційоване з ребром e . Функція $\alpha(n)$ визначає назву поняття, а $\lambda(n)$ — унікальний ідентифікатор вузла орграфа.

Визначимо операції на онтології O .

Унарний оператор визначення кореня онтології *root*. Оскільки коренем онтології O_i є вершина яка не містить дуг, що входять до неї. Відповідно оператор має такий вигляд:

$$root(O_j) = \{n | n \in V, nR_d n_i \ \& \ \neg n_i R_d n\}, \quad (6)$$

де $i = 1, \dots, |O|$, $j = 1, \dots, |P|$. На рис. 1 зафарбований елемент, що є коренем O , відображає результат застосування *root(O)*.

Унарний оператор визначення поняття, що породжує *instanceOf*.

$$instanceOf(n_i) = \{n | n \in V \ \& \ nR_d n_i\}, \quad (7)$$

де $i = 1, \dots, |O|$.

На рис. 2 зафарбований елемент (предок n_5) відображає результат застосування $instanceOf(n_5)$.

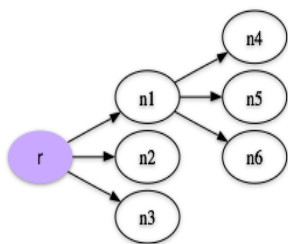


Рис. 1. Графічне зображення результату застосування оператора $root(O_j)$

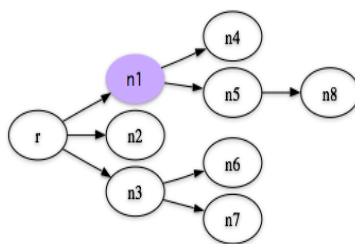


Рис. 2. Графічне зображення результату застосування оператора $instanceOf(n_i)$

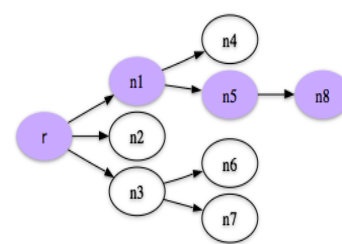


Рис. 3. Графічне зображення результату застосування оператора χ

Оператор знаходження елементів ієрархії заданого поняття $\chi(\alpha, \theta, \lambda)$, де α — власне поняття, для якого необхідно визначити елементи ієрархії, θ — онтологія, в якій необхідно здійснювати пошук, λ — рівень поняття в таксономічній ієрархії θ . Для визначення цього оператора задамо примітивно-рекурсивну функцію

$$\begin{cases} \chi(\alpha, \theta, 0) = root(\theta); \\ \chi(\alpha, \theta, \lambda) = \alpha; \\ \chi(\alpha, \theta, \lambda - 1) = instanceOf(\chi(\alpha, \theta, \lambda)). \end{cases} \quad (8)$$

Результатом застосування оператора $\chi(\alpha, \theta, \lambda)$ є впорядкована множина $(x|x \in V)$, де $\inf(\chi(\alpha, \theta, \lambda)) = \alpha$, а $\sup(\chi(\alpha, \theta, \lambda)) = root(\theta)$. На рис. 3 зафарбовані елементи, що формують ієрархію спадкування для n_8 , відображають результат застосування $\chi(n_8, \theta, 4)$.

Унарний оператор отримання вектора слівформ для поняття онтології $\mu(\alpha(n_i))$. Цей оператор є функціональним відношенням

$$\mu^{-1}(\alpha(V_i)) = \left\{ (m_1, m_2, \dots, m_n) \in U(M) \mid \mu(m_1, m_2, \dots, m_n) = V_i \ \& \ \text{dom}(\mu) = \bigcup_{V_i \in V} \mu^{-1}(V_i) \right\}, \quad (9)$$

де $\alpha(V_i)$ — назва поняття; $U(M)$ — універсальна множина слівформ; (m_1, m_2, \dots, m_n) — множина слівформ для $\alpha(V_i)$.

На рис. 4 зображено приклад результату застосування даного оператора для $\mu(word)$.

Оператор узгодження онтологій на основі експертних правил $\omega(O_i, O_j, A_i, A_j)$. Нехай маємо множину A_i правил узгодження онтологій з суміжних предметних областей. Правила A_i надаються експертами для здійснення точнішого процесу узгодження і задаються у вигляді множини виразів

$$(X \ instanceOf \ O_1.object) \vee (X \ has \ X.option) \vee (O_2 \ hasConcept \ Z) \Rightarrow X \ instanceOf \ O_2.Z, \quad (10)$$

де O_1, O_2 — онтології, що узгоджуються; X, Z — поняття.

Введемо оператор перетворення правил узгодження $\xi(O_1, O_2, A)$ у відношення, яке характеризує результат перетворення, тобто

$$\xi(O_1, O_2, A) \rightarrow (\lambda(O_1.n_j), \lambda(O_2.n_k), \delta) \in R_t, \quad (11)$$

де $O_1.n_j$ та $O_2.n_k$, відповідно, поняття з першої та другої онтологій між якими встановлюється

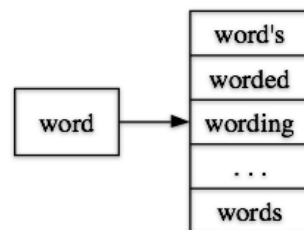


Рис. 4. Графічна інтерпретація результату застосування оператора μ

відношення δ задані у множині правил узгодження A .

Функція λ , як було зазначено вище, визначає унікальні ідентифікатори понять, які необхідні для коректного злиття декількох онтологій.

Оператор узгодження в такому випадку набуде такого вигляду:

$$\omega(O_1, \dots, O_n, A_1, \dots, A_n) = O \left(\bigcup_{i=1}^n V_i, \left(\bigcup_{i=1, j=1, i < j}^n \xi(O_i, O_j, A_i \cup A_j) \right) \cup \left(\bigcup_{i=1}^n R_i \right) \right). \quad (12)$$

Отже результатом узгодження онтологій є нова онтологія, створена за правилами вказаними в (12).

Унарний оператор узгодження онтологій на основі морфологічного словника $\omega_M(O_1, O_2, M)$.

Нехай маємо множину $M = \bigcup_{i=1}^{|V|} \mu^{-1}(\alpha(V_i))$, що містить словоформи для всіх понять інформаційної статті. Введемо оператор перетворення правил узгодження $\xi_M(O_1, O_2, M)$ по аналогії з (11)

$$\xi_M(O_1, O_2, M) \rightarrow (\lambda(O_1.n_j), \lambda(O_2.n_k), isA) \in R_{tm}, \quad (13)$$

де isA — відношення еквівалентності, що встановлюється між елементами $O_1.n_j$ та $O_1.n_k$ онтологій O_1 та O_2 .

Множина понять для морфологічного узгодження визначається таким чином: $S_m = (n | n \subseteq V_1 \ \& \ \mu^{-1}(\alpha(n)) \subseteq V_2)$ Оператор морфологічного узгодження в цьому випадку набуде такого вигляду:

$$\omega_M(O_1, \dots, O_n, M) = O \left(\bigcup_{i=1}^n V_i, \left(\bigcup_{i=1, j=1, i < j}^n \xi_M(O_i, O_j, M) \right) \cup \left(\bigcup_{i=1}^n R_i \right) \right). \quad (14)$$

Оператор узгодження на основі експертних правил та морфологічного аналізу буде мати вигляд

$$\begin{aligned} \omega_G(O_1, \dots, O_n, A_1, \dots, A_n, M) = \\ = O \left(\bigcup_{i=1}^n V_i, \left(\bigcup_{i=1, j=1, i < j}^n \xi(O_i, O_j, A_i \cup A_j) \right) \cup \left(\bigcup_{i=1, j=1, i < j}^n \xi_M(O_i, O_j, M) \right) \cup \left(\bigcup_{i=1}^n R_i \right) \right). \end{aligned} \quad (15)$$

Розглянуті операції формують багатоосновну алгебру $G_{ont} = (U, \Omega)$, де комплект задано, як $U = \{O_i, A, M\}$, а множину багатосортних операцій, як

$$\Omega = \{root, instance\ Of, \chi, \xi, \xi_M, \omega, \omega_M, \omega_G\}.$$

Розглянемо застосування запропонованої алгебри на прикладі онтологій трьох Інтернет-ресурсів — сайту Державного управління екології та природних ресурсів Вінницької області (ДУ-ЕПРВО (SDERVR)) (vineco.gov.ua), онтологія якої зображена на рис. 5, системи моніторингу поверхневих вод Вінницької області (СМПВВО (MSSWVR)) (edem.vstu.vinnica.ua/monitoring), онтологія якої зображена на рис. 6, та Вінницького національного технічного університету (ВНТУ (VSTU)) (vntu.edu.ua), онтологія якої зображена на рис. 7. Зазначені ресурси є логічно пов'язаними, оскільки всі мають відношення до ВНТУ, але їх гіпертекстова структура не містить перехресних посилань.

Необхідно зазначити, що онтології на рис. 5—7 неповні, вони лише відображають частини, що використовуються для ілюстрації процесу та результату узгодження. Візуалізація онтологій здійснена у пакеті програм створення онтологій Protege [4].

Перелік експертних правил узгодження для вищезазначених Інтернет-ресурсів такий:

vineco.gov.ua:

SDERVR isA ДУЕПРВО

ДУЕПРВО isA Державне управління екології та природних ресурсів Вінницької області;

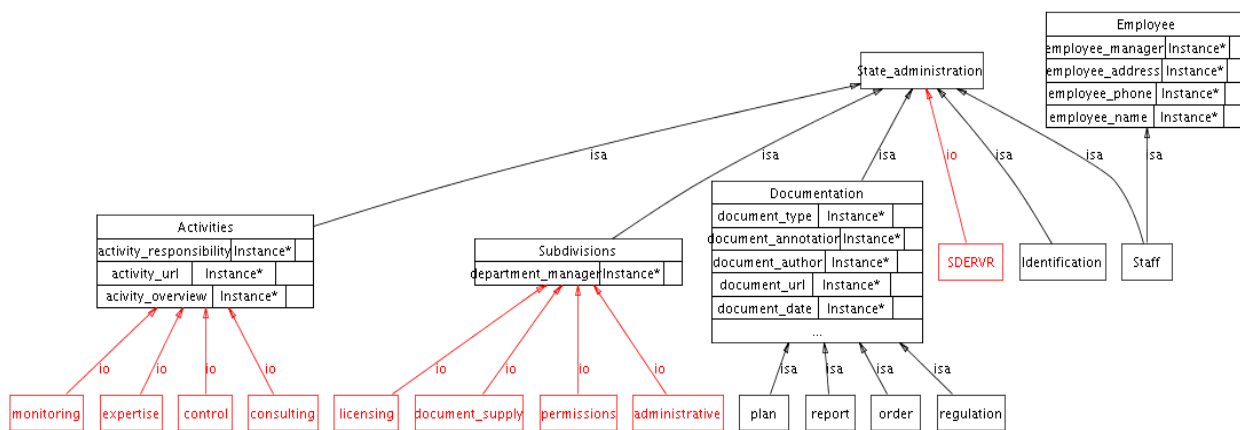


Рис. 5. Графічне зображення онтології ресурсу vinесо.gov.ua

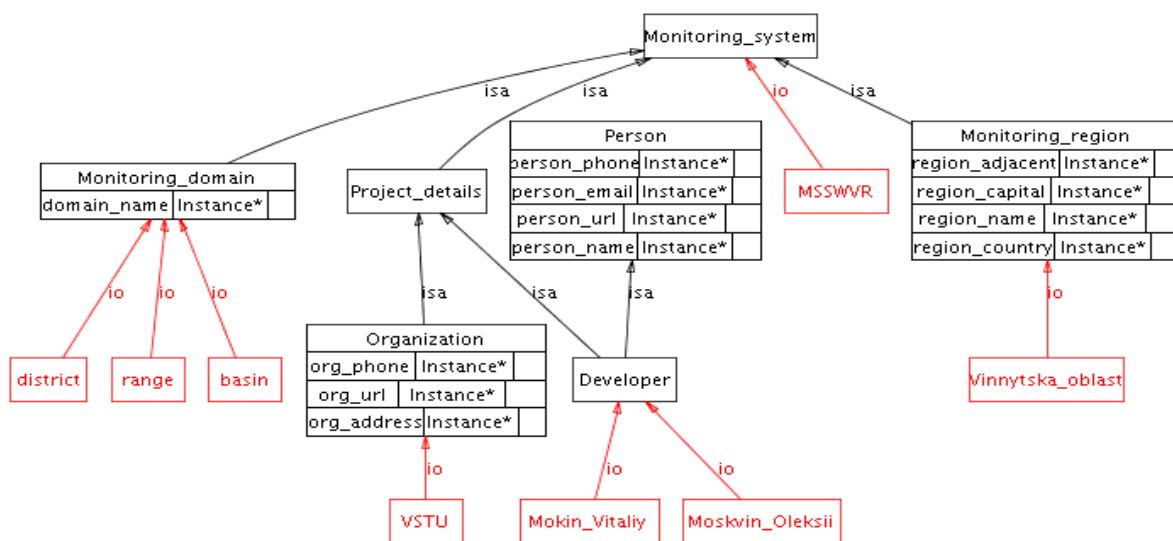


Рис. 6. Графічне зображення онтології ресурсу edem.vstu.vinnica.ua/monitoring

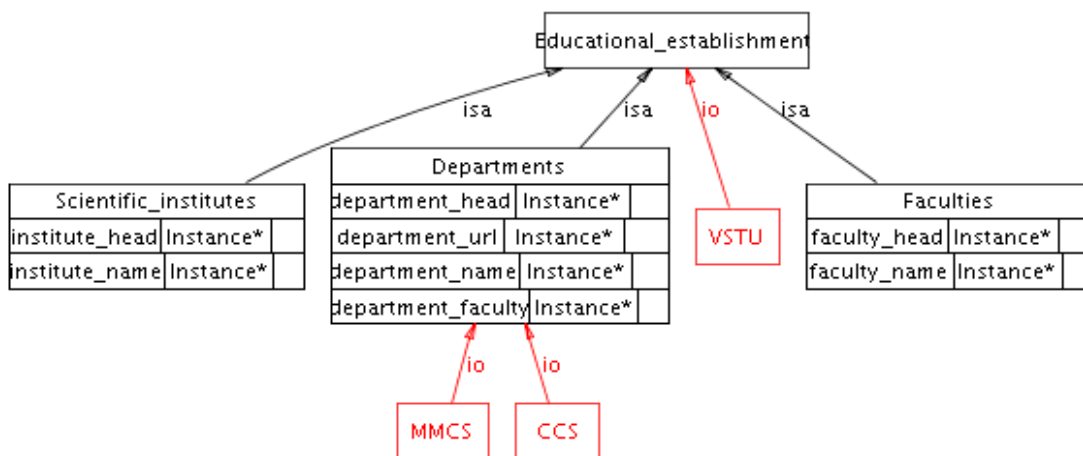


Рис. 7. Графічне зображення онтології ресурсу vntu.edu.ua

edem.vstu.vinnica.ua/monitoring:

MSSWVR isA СМІПВВО

СМІПВВО isA Система моніторингу поверхневих вод Вінницької області

СМІПВВО isA Моніторинг

VSTU isA BHTV

BHTV isA Вінницький національний технічний університет;

yntu.edu.ua:

MMCS isA MMCC

MMCC isA Кафедра моделювання та моніторингу складних систем.

Виконаємо узгодження для трьох вищезазначених ресурсів. Оператор узгодження на основі експертних правил, визначений у (12), для цього випадку має такий вигляд:

$$\omega(O_1, O_2, O_3, A_1, A_2, A_3) = \\ = O(V_1 \cup V_2 \cup V_3, (\xi(O_1, O_2) \cup \xi(O_2, O_3) \cup \xi(O_1, O_3)) \cup (A_1 \cup A_2 \cup A_3) \cup (R_1 \cup R_2 \cup R_3)),$$

де $V_1 = \{\text{State administration, SDERVR, activities, subdivisions, documentation, identification, staff, monitoring, expertise, control, consulting, licensing, document supply, permissions, administrative, plan, report, order, regulation}\}$; $V_2 = \{\text{Monitoring system, MSSWVR, monitoring domain, project details, organization, developer, monitoring region, district, range, basin, VSTU, Mokin Vitaliy, Moskvin Oleksii, Vinnytska oblast}\}$;

$V_3 = \{\text{Educational establishment, VSTU, Scientific institutes, departments, faculties, MMCS, CCS}\}$;

$\xi(O_1, O_3) = \emptyset$;

$\xi(O_1, O_2) = \{(O_1.\text{monitoring}, O_2.\text{MSSWVR}, \text{isA})\}$, $\xi(O_2, O_3) = \{(O_2.\text{VSTU}, O_3.\text{VSTU}, \text{isA})\}$.

Результат узгодження зображено на рис. 8. Елементи таксономій, між якими були встановлені зв'язки, показані у прямокутниках. Відповідно зв'язки між онтологіями показані пунктирними лініями.

Отже, в роботі запропоновано математична формалізація операцій для узгодження онтологічних контекстів, що на відміну від існуючих включає формалізовані засоби морфологічного узгодження та операції для ідентифікації ієрархій. Тестові приклади показують коректні результати у разі використання тематично пов'язаних онтологій.

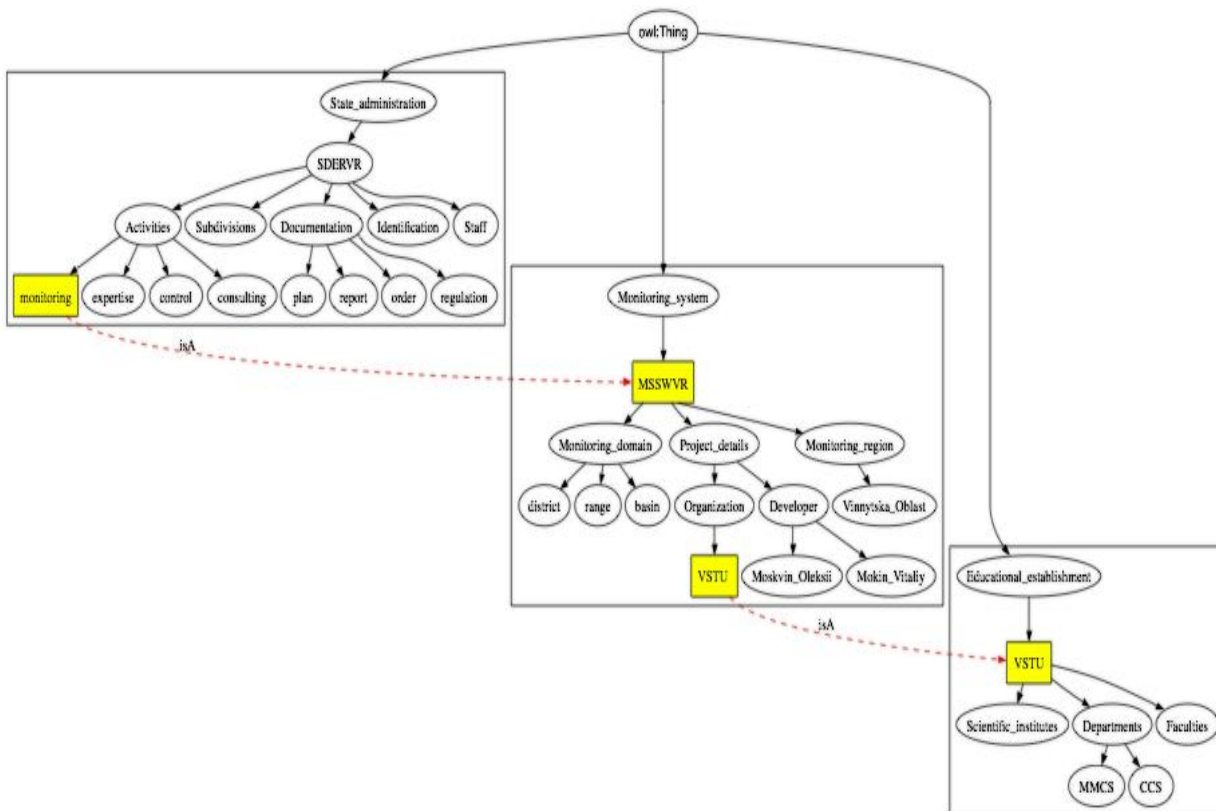


Рис. 8. Результат узгодження трьох онтологій

Висновки

Запропонована алгебра операцій для маніпулювання онтологіями, яка необхідна для формалізації операцій узгодження та пошуку елементів в ієрархіях їх таксономій в тематично пов'язаних областях знань. Ця алгебра дозволяє сформуванню базис для опису більш складних операцій на онтологіях та їх складових, таких як оцінка ваги понять в ієрархіях таксономій, а також дозволяє автоматизувати цей процес.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Mitra P. An Ontology Composition Algebra / P. Mitra, G. Wiederhold // Handbook on Ontologies; eds. by S. Staab, R. Studer. Berlin; Heidelberg; New York : Springer-Verlag, 2003. P. 93—113.
2. Теория информации / [Ю. В. Капітонова, С. Л. Кривий, О. А. Летичевський и др.] — К. : «Наукова Думка», 2002. — 578 с.
3. Steffen Staab, Rudi Studer (Eds.) // Handbook on Ontologies, Springer Verlag, Heidelberg; New York : Springer-Verlag, 2003.
4. A Practical Guide To Building OWL Ontologies Using Protege 4 and CO-ODE Tools [Електронний ресурс] / Matthew Horridge // The University Of Manchester. — Mar. 24, 2011 — Режим доступу до журналу : http://owl.cs.manchester.ac.uk/tutorials/protegeowltutorial/resources/ProtegeOWLTutorialP4_v1_3.pdf.

Рекомендована кафедрою комп'ютерних систем управління

Стаття надійшла до редакції 22.07.11
Рекомендована до друку 5.09.11

Дубовой Володимир Михайлович — завідувач кафедри, *Москвін Олексій Михайлович* — аспірант.
Кафедра комп'ютерних систем управління, Вінницький національний технічний університет, Вінниця